



Deficits in audiovisual speech perception in normal aging emerge at the level of whole-word recognition



Ryan A. Stevenson^{a,b,c,d,*}, Caitlin E. Nelms^{e,f}, Sarah H. Baum^{c,g}, Lilia Zurkovsky^h,
Morgan D. Barense^{a,i}, Paul A. Newhouse^h, Mark T. Wallace^{b,c,d,h,j}

^a Department of Psychology, University of Toronto, Toronto, Ontario, Canada

^b Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

^c Vanderbilt Brain Institute, Nashville, TN, USA

^d Vanderbilt Kennedy Center, Nashville, TN, USA

^e Department of Psychology, Austin Peay State University, Clarksville, TN, USA

^f Department of Communication Sciences and Disorders, University of Memphis, Memphis, TN, USA

^g Department of Neurobiology and Anatomy, University of Texas Medical School at Houston, TX, USA

^h Center for Cognitive Medicine, Department of Psychiatry, Vanderbilt University, Nashville, TN, USA

ⁱ Rotman Research Institute, Toronto, Ontario, Canada

^j Department of Psychology, Vanderbilt University, Nashville, TN, USA

ARTICLE INFO

Article history:

Received 20 April 2014

Received in revised form 22 July 2014

Accepted 2 August 2014

Available online 7 August 2014

Keywords:

Speech perception

Multisensory

Aging

Multisensory integration

Inverse effectiveness

ABSTRACT

Over the next 2 decades, a dramatic shift in the demographics of society will take place, with a rapid growth in the population of older adults. One of the most common complaints with healthy aging is a decreased ability to successfully perceive speech, particularly in noisy environments. In such noisy environments, the presence of visual speech cues (i.e., lip movements) provide striking benefits for speech perception and comprehension, but previous research suggests that older adults gain less from such audiovisual integration than their younger peers. To determine at what processing level these behavioral differences arise in healthy-aging populations, we administered a speech-in-noise task to younger and older adults. We compared the perceptual benefits of having speech information available in both the auditory and visual modalities and examined both phoneme and whole-word recognition across varying levels of signal-to-noise ratio. For whole-word recognition, older adults relative to younger adults showed greater multisensory gains at intermediate SNRs but reduced benefit at low SNRs. By contrast, at the phoneme level both younger and older adults showed approximately equivalent increases in multisensory gain as signal-to-noise ratio decreased. Collectively, the results provide important insights into both the similarities and differences in how older and younger adults integrate auditory and visual speech cues in noisy environments and help explain some of the conflicting findings in previous studies of multisensory speech perception in healthy aging. These novel findings suggest that audiovisual processing is intact at more elementary levels of speech perception in healthy-aging populations and that deficits begin to emerge only at the more complex word-recognition level of speech signals.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Visual cues are known to significantly impact speech perception; when one can both hear a speaker's utterance and concurrently see the articulation of that utterance (lip reading), speech comprehension is more accurate (Ross et al., 2007a, 2011; Sommers et al., 2005; Stevenson and James, 2009; Sumby and Pollack, 1954)

and less effortful (Fraser et al., 2010) than when only auditory information is available. The behavioral gain observed when processing information via multiple sensory modalities is governed by a number of factors, with one of the most important being the relative effectiveness of the stimuli that are paired. As a general rule, greater benefits are observed from pairing stimuli that, on their own, are each weakly effective, compared with pairing stimuli that are both strongly effective when presented in isolation (Bishop and Miller, 2009; James et al., 2012; Meredith and Stein, 1983, 1986; Nath and Beauchamp, 2011; Stevenson and James, 2009; Stevenson et al., 2007, 2009, 2012a, 2012b; Wallace et al., 1996; Werner and Noppeney, 2009). This concept of “inverse effectiveness” implies

* Corresponding author at: Department of Psychology, University of Toronto, Rm 523, Sidney Smith, 100 St. George Street, Toronto, ON, Canada M6G 3G3. Tel.: +1 416 978 5464; fax: +1 416 978 4811.

E-mail address: ryan.andrew.stevenson@gmail.com (R.A. Stevenson).

that the primary benefits of multisensory integration take place when the individual stimuli provide weak or ambiguous information. For example, the addition of a visual speech signal provides the greatest gain when the auditory speech is noisy (Sumbly and Pollack, 1954). Once an individual stimulus is sufficiently salient, the need for multisensory-mediated benefits substantially declines.

In the context of real-world multisensory stimuli, these changes in effectiveness can be mediated not only by changes in the external characteristics of the stimuli (e.g., the loudness of an auditory stimulus) but also by changes in internal events governing the processing of that information. The declines in visual and auditory acuity associated with normal aging are probably the result of decreases in internal signal strength attributable to changes in transduction and encoding processes but also by additional internal noise (i.e., variability) to the transduction and encoding processes. The loss of visual and auditory acuity is seen for both simple and more complex stimuli and is particularly prevalent for speech signals, most notably in the presence of external noise (Dubno et al., 1984; Gosselin and Gagne, 2011; Humes, 1996; Martin and Jerger, 2005; Sommers et al., 2005). Although these age-related declines in speech perception and comprehension have been widely interpreted to be a result of changes in auditory acuity (Liu and Yan, 2007) and diminished ability to filter task-irrelevant auditory information (Hugenschmidt et al., 2009), declines in visual acuity may play an important and underappreciated role. Some evidence suggests that older adults may rely on visual information to a greater extent than their younger counterparts (Freiherr et al., 2013; Laurienti et al., 2006), which may reflect the use of multisensory integration as a compensatory mechanism for declining unisensory abilities.

In this prior work, multisensory gain increased with age for the integration of simple audiovisual stimuli such as flashes and tones, a finding consistent with the principle of inverse effectiveness given the age-related declines in unisensory processing acuity (Laurienti et al., 2006). However, for speech-related stimuli, the picture is more complex and provides only partial support for the concept of inverse effectiveness. Whereas older and younger adults showed equivalent levels of audiovisual gain for high (i.e., easier) signal-to-noise ratio (SNR) trials, older adults showed less gain than younger adults on low (i.e., more difficult) SNR trials (Tye-Murray et al., 2010). One potential explanation for these disparate findings is that in this latter study participants were required to complete or repeat whole sentences, which may introduce variability because of other cognitive factors. For example, verbal memory is known to decline in nondemented aging (Park et al., 2002), a finding that may impair the ability of older participants to recall whole sentences. Thus, the reduced multisensory gain observed on low SNR trials in older adults relative to younger adults may in fact reflect memory impairments for the full sentences, rather than deficits in the integration of auditory and visual cues.

Here, we conducted a novel study designed to address these conflicting observations, specifically structured to examine how aging affects multisensory-mediated gains in speech perception under noisy conditions. Critically, we examined these gains at the level of more elementary (i.e., phonemic) and more complex (i.e., whole word) components of speech, providing the first systematic investigation of how multisensory integration at different levels of processing is affected by aging. We presented younger and older healthy adults with a standard audiovisual speech-in-noise task in which participants reported the perceived word. We used single-word presentations to limit the impact of higher-order cognitive changes known to occur with aging, such as changes in memory or context. Importantly, the task was scored both at the whole-word level and the phoneme level, allowing us to pinpoint whether changes in multisensory gain across the life span differed depending on the level of processing necessary for accurate comprehension.

Our results provide evidence that older adults show largely intact multisensory processing at lower (i.e., phonemic) levels of speech perception but begin to show deficits at higher level processing with whole-word recognition.

2. Methods

2.1. Participants

Thirty-four participants (20 females, mean age = 39.0 years, SD = 18.4, range = 19–67 years) completed a behavioral speech-in-noise paradigm. Experimental protocols were approved by Vanderbilt University Medical Center's institutional review board. Participants were divided into 2 age groups based on the median age (39 years) reported in Ross et al. (2007a) publication investigating inverse effectiveness in single-word recognition. The younger group included participants 39 years of age or younger (18 in total, 8 females, mean age = 22.8 years, SD = 4.7, range = 19–38 years), and the older group consisted of participants 40 years of age or older (16 in total, 12 females, mean age = 57.3 years, SD = 6.9, range = 45–67 years). For demographic information, see Table 1. All individuals were screened for normal visual acuity with a tumbling E visual chart and were not hearing-aid users. Additionally, Mini-mental state examinations were conducted on all participants, with a score greater than or equal to 27 used as an exclusionary cutoff, though no participants were excluded. Additionally, participants reported no neurologic impairments. Participants were recruited via flyer at Vanderbilt University Medical Center.

2.2. Stimuli

Stimuli included dynamic, audiovisual (AV) recordings of a female speaker saying 216 tri-phonemic nouns. Stimuli were selected from a previously published stimulus set, The Hoosier Audiovisual Multi-talker Database (Sheffert et al., 1996). All stimuli were spoken by speaker F1. The stimuli selected were monosyllabic English words that were matched across sets for accuracy on both visual-only and audio-only recognition (Lachs and Hernandez, 1998) and were also matched across sets in lexical neighborhood density (Luce and Pisoni, 1998; Sheffert et al., 1996). This set of single-word tokens have been used successfully in previous studies of multisensory integration (Stevenson and James, 2009; Stevenson et al., 2009, 2011). Audio signal levels were measured as root mean square contrast and equated across all tokens.

All stimuli throughout the study were presented using MATLAB 2012b (MATHWORKS Inc, Natick, MA, USA) software with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Visual stimuli were 200 × 200 pixels and subtended 10 × 10° of visual angle. Audio stimuli were presented through 2 aligned speakers on each side of the monitor. All tokens lasted 2 seconds and included all prearticulatory gestures.

In the visual-only condition, the visual component of each stimulus or viseme, was presented. Auditory stimuli were all overlaid with 8-channel multitalker babble at 72 dB SPL. The presentation of auditory babble presentation began 500 ms before the beginning of the stimulus token and ended 500 ms following token offset. The root mean square of the auditory babble was linearly ramped up and

Table 1
Participant demographics

	N	Mean age (SD)	Age range (y)	Female (%)	Male (%)
Younger	18	22.8 (4.7)	19–38	44	56
Older	16	57.3 (6.9)	45–67	75	25

Key: SD, standard deviation.

down, respectively, during the pre- and post-stimulus 500 ms periods, and was presented with the first and last frames of the visual token, respectively. Auditory stimuli were presented at 4 separate sound levels relative to the auditory noise. Differences in auditory level, or SNR, included 0, –6, –12, and –18 dB SPL.

2.3. Procedures

Participants sat in a sound- and light-attenuating WhisperRoom (Model SE 2000; Whisper Room Inc) approximately 60 cm from the monitor. Participants were presented with 9 separate runs of 24 single-word presentations each: 4 audiovisual runs (one at each SNR), 4 auditory-only runs (one at each SNR), and 1 visual-only run (with auditory multitalker babble). During auditory-only presentations, the first frame of the associated video was presented and remained static throughout the presentations. Run orders were randomized across participants. Within participants, word lists were randomized across runs with no words repeated. Word lists were also counterbalanced between individuals, so words were presented in different modalities and SNRs for each individual.

Experimental procedures were identical for all runs. Participants were instructed to attend to the speaker at all times, and to report the word they perceived by typing the word (on a keyboard placed in front of them). After each trial, the experimenter confirmed the participant's report to correct for spelling errors, and then the next word was presented. No time limit was given for participant responses. Each run lasted approximately 5 minutes, and all run orders were counterbalanced.

2.4. Analysis

Responses were scored in 2 ways, at the whole-word level and at the phoneme level. Word-recognition accuracy was scored as correct if and only if the entire word reported was correctly perceived. Phoneme accuracy allowed for participants to be scored as correctly reporting 0–3 phonemes per word. Mean word accuracy and phoneme accuracy were then calculated for each participant and for each run.

The expected multisensory accuracy predicted by the individual unisensory responses was calculated by:

$$pAV = p(A) + p(V) - [p(A) \times p(V)]$$

where pAV represents a null hypothesis characterizing what the response will be to audiovisual presentations if the auditory and visual information are processed independently (Stevenson et al.,

2014a), and where $p(A)$ and $p(V)$ represent the individual's response accuracy to auditory- and visual-only presentations, respectively. Each participant's responses to unisensory presentations were used to calculate their individual predicted audiovisual (pAV) values for each SNR level. These predicted values were then used as a null hypothesis from which we measured multisensory interactions, namely multisensory gain which we define here as an increase in performance above and beyond that predicted by nonintegrative statistical facilitation (Raab, 1962).

To assess unisensory auditory sensory processing levels, a 2 (age group: younger vs. older) \times 4 (SNR: 0, –6, –12, and –18 dB SPL), mixed-model analysis of variance (ANOVA) was conducted for both at the word- and phoneme-recognition levels for auditory-only word presentations. To assess unisensory visual processing levels (i.e., lip reading), a between-subject t test was performed across age groups. To assess multisensory benefits, a 2 (age group: younger vs. older) \times 4 (SNR: 0, –6, –12, and –18 dB SPL) \times 2 (AV measure: observed AV vs. predicted AV) mixed-model ANOVA was conducted for both word and phoneme accuracy scores. Finally, to directly compare age-related differences in multisensory gain for word versus phoneme recognition, the percent multisensory gain data were subjected to a 2 (age group: younger vs. older) \times 4 (SNR: 0, –6, –12, and –18 dB SPL) \times 2 (measure: whole-word vs. phoneme) mixed-model ANOVA.

3. Results

All analyses were run separately for each of the 2 levels of data scoring: whole-word accuracy and phoneme accuracy. Mean accuracies and standard deviations were calculated for each group, sensory modality, and SNR level (Table 2).

3.1. Unisensory performance

3.1.1. Visual-only whole-word and viseme recognitions

A between-subject t test was used to compare younger and older groups' visual-only speech perception accuracies at both the whole-word and viseme recognition levels. In both cases, younger individuals showed higher rates of visual-only accuracy. These differences approached significance at the whole-word recognition level (Fig. 1A) and at the viseme recognition level (Fig. 2A). For statistical results, see Table 2.

3.1.2. Auditory-only whole-word recognition

To analyze accuracies with auditory-only presentations, a 2-way, mixed-method, repeated-measures ANOVA was conducted at both

Table 2
Mean whole-word and phoneme-recognition accuracy rates

Modality	SNR	Word recognition					Phoneme recognition				
		Younger group	Older group	<i>p</i>	<i>t</i>	<i>d</i>	Younger group	Older group	<i>p</i>	<i>t</i>	<i>d</i>
V	N/A	14.1 \pm 8.7	9.4 \pm 7.4	0.104	1.67	0.59	43.1 \pm 10.6	37.5 \pm 12.3	0.135	1.54	0.49
A	0	86.1 \pm 9.4	71.8 \pm 15.7	0.002	3.45	1.11	94.2 \pm 4.2	56.9 \pm 9.8	0.005	3.00	0.97
	–6	73.4 \pm 9.2	52.6 \pm 15.1	<0.001	4.42	1.67	88.1 \pm 4.7	75.3 \pm 10.5	<0.001	4.86	1.57
	–12	58.1 \pm 7.5	38.5 \pm 14.8	<0.001	4.113	1.67	76.7 \pm 6.0	62.6 \pm 12.2	<0.001	4.39	1.47
	–18	7.2 \pm 6.7	2.9 \pm 4.7	0.062	1.94	0.74	23.7 \pm 10.0	14.6 \pm 9.1	0.008	2.86	0.95
AV	0	92.8 \pm 7.5	80.7 \pm 23.5	0.084	1.79	0.69	97.1 \pm 2.9	89.2 \pm 23.4	0.154	1.46	0.48
	–6	86.8 \pm 8.2	77.1 \pm 13.8	0.048	2.06	0.86	95.1 \pm 3.5	90.4 \pm 6.1	0.008	2.85	0.95
	–12	78.0 \pm 12.0	63.5 \pm 16.3	0.001	3.54	1.01	89.2 \pm 7.5	83.0 \pm 8.5	0.022	2.41	0.77
	–18	45.6 \pm 13.0	27.1 \pm 12.4	<0.001	4.80	1.46	68.7 \pm 8.2	57.4 \pm 11.9	0.002	3.39	1.10
Multisensory gain ^a	0	5.2 \pm 11.2	6.90 \pm 15.9	0.447	0.77	0.12	0.6 \pm 3.9	–2.0 \pm 17.7	0.548	0.61	0.20
	–6	9.9 \pm 8.9	20.0 \pm 12.2	0.006	2.90	0.94	2.0 \pm 3.7	6.1 \pm 6.5	0.021	2.43	0.78
	–12	14.3 \pm 11.1	19.2 \pm 14.3	0.818	0.23	0.39	2.7 \pm 7.1	7.2 \pm 7.6	0.098	1.71	0.62
	–18	25.5 \pm 10.8	15.0 \pm 10.4	0.001	3.56	0.99	12.4 \pm 9.4	10.7 \pm 12.4	0.635	0.48	0.16

p, *t*, and *d* values describe a comparison of accuracies between younger and older groups.

Key: A, auditory; AV, audiovisual; SNR, signal-to-noise ratio; V, visual.

^a Values given for multisensory gain are percent changes between observed and predicted recognition accuracy rates.

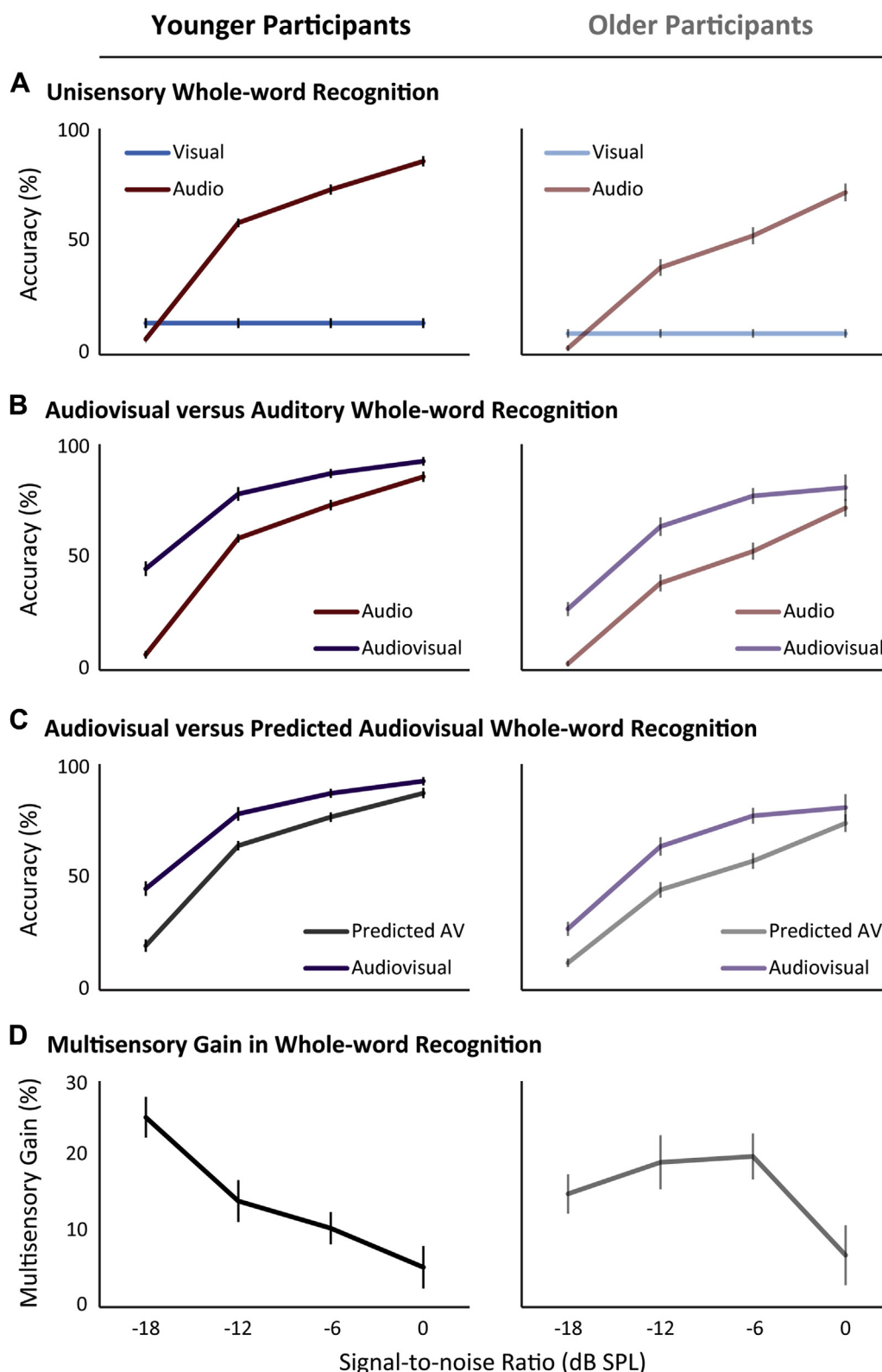


Fig. 1. Word recognition. For all panels, younger adults are depicted on the left and older adults on the right. Panel A depicts word-recognition accuracies with auditory-only across signal-to-noise ratios (SNR) and visual-only presentations. Panel B compares similar accuracies with audiovisual presentations to the respective auditory-only conditions. Panel C compares the accuracies to audiovisual presentations with the predicted accuracy level based on the unisensory responses with the assumption that there are no interactions between auditory and visual processing. Finally, panel D shows the level of behavioral multisensory gain when stimuli are presented concurrently through 2 stimulus modalities instead of the level predicted by independent presentations of the same stimuli calculated as $AV - p(AV)$. Error bars depict standard error. Abbreviations: AV, audiovisual; pAV, predicted audiovisual.

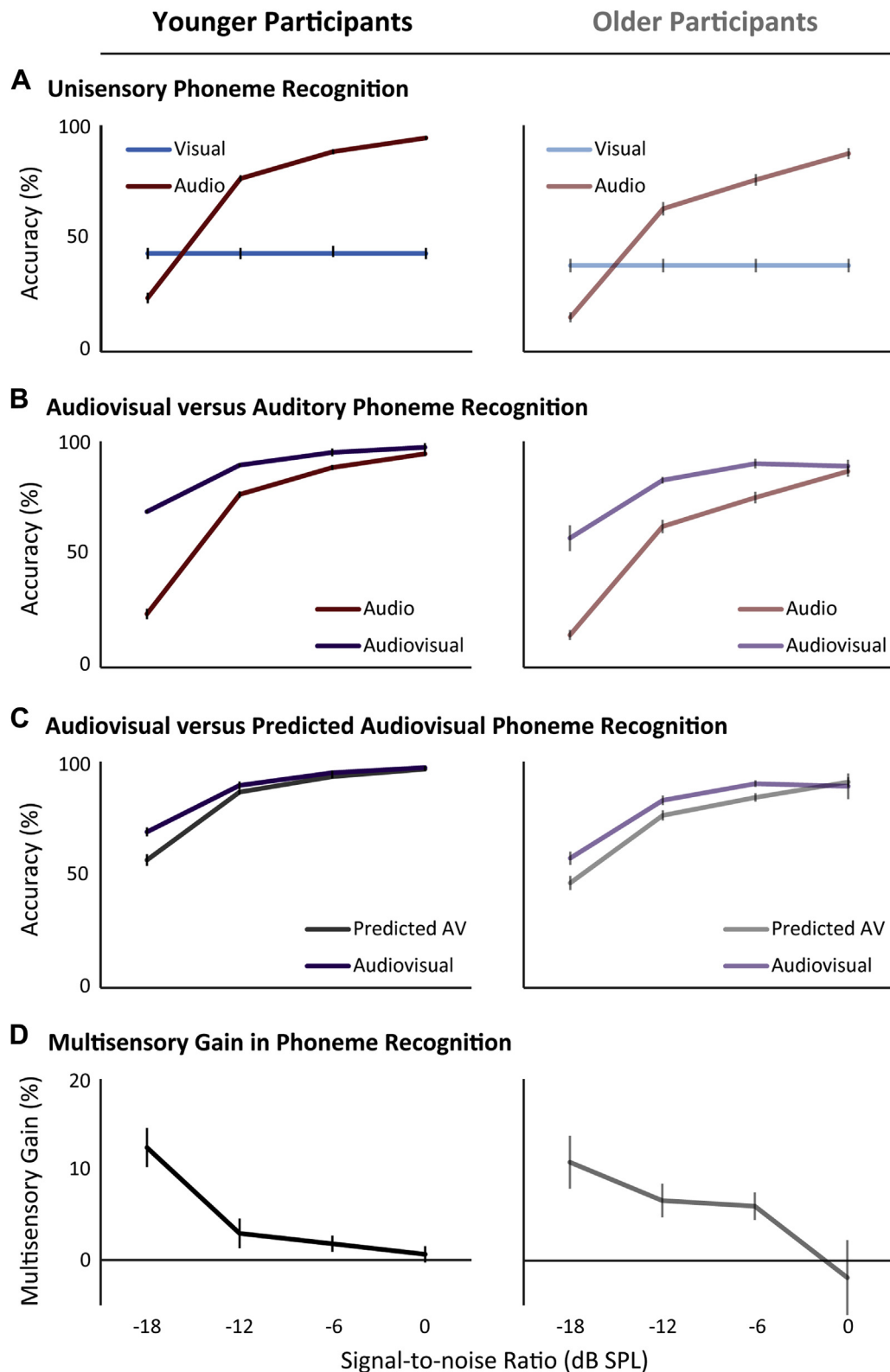


Fig. 2. Phoneme recognition. For all panels, younger adults are depicted on the left and older adults on the right. Panel A depicts phoneme-recognition accuracies with auditory-only and visual-only presentations across signal-to-noise ratios (SNR). Panel B compares similar accuracies with audiovisual presentations to the respective auditory-only conditions. Panel C compares the accuracies to audiovisual presentations with the predicted accuracy level based on the unisensory responses with the assumption that there are no interactions between auditory and visual processing. Finally, panel D shows the level of behavioral multisensory gain when stimuli are presented concurrently through 2 stimulus modalities instead of the level predicted by independent presentations of the same stimuli calculated as $AV - p(AV)$. Error bars depict standard error. Abbreviations: AV, audiovisual; pAV, predicted audiovisual.

the whole-word and phoneme-recognition levels with age group as a between-subjects factor and SNR as a within-subjects factor. For whole-word recognition (Fig. 1B), a main effect of age group was seen, with younger adults showing higher accuracy than older adults ($p < 0.001$, $F_{(1,33)} = 27.86$, $\text{partial-}\eta^2 = 0.47$). A main effect of SNR was also observed, with higher SNRs resulting in higher accuracy rates ($p < 0.001$, $F_{(3,30)} = 552.12$, $\text{partial-}\eta^2 = 0.98$). Finally, a significant interaction between age group and SNR levels was observed, with older adults showing decreases in accuracy at higher SNR levels as compared with younger adults ($p = 0.001$, $F_{(3,30)} = 7.80$, $\text{partial-}\eta^2 = 0.44$). For statistical results, see Table 2.

3.1.3. Auditory-only phoneme recognition

For phoneme recognition with auditory-only presentations (Fig. 2B), a main effect of age group was also seen, with younger adults showing higher accuracy than older adults ($p < 0.001$, $F_{(1,33)} = 24.83$, $\text{partial-}\eta^2 = 0.43$). A main effect of SNR was also seen for phoneme recognition, with higher SNRs leading to higher accuracy rates ($p < 0.001$, $F_{(3,30)} = 558.70$, $\text{partial-}\eta^2 = 0.98$). Finally, no interaction between age group and SNR levels was observed ($p = 0.16$, $F_{(3,31)} = 1.87$, $\text{partial-}\eta^2 = 0.15$). For statistical results, see Table 2.

3.2. Multisensory performance

3.2.1. Multisensory whole-word recognition

For the analysis of the effects of combined congruent audiovisual presentations, comparisons were made between the observed AV response and the pAV response based on a given individual's unisensory accuracies. A 3-way, repeated-measures ANOVA was conducted on whole-word recognition scores with age group as a between-subject factor, and SNR level and observed versus predicted multisensory accuracy (AV and pAV) as within-subjects factors. A significant main effect of age group was observed with the younger adults showing higher whole-word recognition accuracy (Fig. 1C, $p < 0.001$, $F_{(1)} = 21.50$, $\text{partial-}\eta^2 = 0.40$). A significant main effect of SNR was also observed, with higher SNRs associated with more accurate whole-word recognition responses (Fig. 1C, $p < 0.001$, $F_{(3)} = 392.55$, $\text{partial-}\eta^2 = 0.98$).

Additionally, a main effect of observed accuracy versus predicted accuracy for these multisensory conditions was observed, with observed audiovisual responses being more accurate than predicted audiovisual whole-word recognition accuracy levels (Fig. 1C, $p < 0.001$, $F_{(1)} = 171.18$, $\text{partial-}\eta^2 = 0.84$). No 2-way interactions were observed between age group and observed versus predicted multisensory accuracy ($p = 0.49$, $F_{(1,33)} = 0.49$, $\text{partial-}\eta^2 = 0.02$) or between age group and SNR level ($p = 0.58$, $F_{(3,30)} = 0.66$, $\text{partial-}\eta^2 = 0.06$). However, a significant interaction was observed between observed versus predicted multisensory accuracy and SNR, where a larger difference was seen between observed AV response accuracy and pAV response accuracy at lower SNRs (Fig. 1D, $p < 0.001$, $F_{(3,30)} = 7.91$, $\text{partial-}\eta^2 = 0.44$). This finding is concordant with the principle of inverse effectiveness, in which greater multisensory gain is seen with decreasing SNR (inverse effectiveness). Finally, a significant 3-way interaction was observed, such that the increasing difference between observed audiovisual response accuracy and predicted accuracy (multisensory gain) was more prominent in the younger group relative to the older group (Fig. 1D, $p = 0.001$, $F_{(3,30)} = 6.90$, $\text{partial-}\eta^2 = 0.41$).

3.2.2. Multisensory phoneme recognition

A 3-way, repeated-measures ANOVA was also conducted for phoneme recognition with age group as a between-subject factor, and SNR level and observed versus predicted multisensory (AV and pAV) as within-subjects factors. A significant main effect of age

group was observed with the younger adults showing higher phoneme-recognition accuracy (Fig. 2B, $p = 0.001$, $F_{(1)} = 12.56$, $\text{partial-}\eta^2 = 0.28$). A significant main effect of SNR was also observed, with higher SNRs associated with more accurate phoneme-recognition responses ($p < 0.001$, $F_{(3)} = 315.02$, $\text{partial-}\eta^2 = 0.97$). For statistical results, see Table 2.

Additionally, a main effect of observed versus predicted multisensory accuracy was observed, with observed audiovisual responses being more accurate than predicted audiovisual phoneme-recognition accuracy responses (Fig. 2C, $p < 0.001$, $F_{(1)} = 62.55$, $\text{partial-}\eta^2 = 0.66$). A significant interaction was observed between observed versus predicted multisensory accuracy and SNR, where a larger difference was seen between AV and pAV phoneme-recognition accuracies at lower SNRs, again in support of inverse effectiveness (Fig. 2C, $p = 0.002$, $F_{(3,31)} = 6.50$, $\text{partial-}\eta^2 = 0.39$). No 2-way interactions were observed between age group and observed versus predicted multisensory performance ($p = 0.44$, $F_{(1,33)} = 0.49$, $\text{partial-}\eta^2 = 0.02$) or between age group and SNR level ($p = 0.43$, $F_{(3,31)} = 0.95$, $\text{partial-}\eta^2 = 0.09$). By contrast to whole-word recognition, no significant 2-way interaction was observed in the multisensory gain (difference between audiovisual and pAV phoneme-recognition accuracies at each SNR level) and age group (Fig. 2D, $p = 0.22$, $F_{(3,31)} = 1.58$, $\text{partial-}\eta^2 = 0.13$).

3.2.3. Age differences in multisensory gain for whole-word and phoneme recognition

As reported previously, measurements of multisensory gain revealed a significant 2-way interaction between age group and SNR for word recognition but not for phoneme recognition. To directly compare age-related multisensory gains in word versus phoneme recognition, we conducted a 2 (age group: younger vs. older) \times 4 (SNR: 0, –6, –12, and –18 dB SPL) \times 2 (measure: whole-word vs. phoneme) ANOVA on the percentage multisensory gain data shown Figs. 1D and 2D. Of primary importance, we observed a significant 3-way interaction ($p = 0.002$, $F_{(3,96)} = 1.58$, $\text{partial-}\eta^2 = 0.14$), indicating that older and younger adults showed significantly different patterns of inverse effectiveness across whole-word and phoneme accuracies. Follow-up *t*-tests were conducted to examine the driving factor in this interaction (for detailed statistics, see Table 2). In brief, these tests revealed that older adults showed significantly more gain at the –6 dB SPL SNR for both whole-word and phoneme recognition. By contrast, at the –18 dB SPL SNR older adults showed significantly reduced gain for whole-word, but not for phoneme, recognition. Thus, only at the level of whole-word recognition did older adults show decreased multisensory gain. Secondly, the expected within-subject main effect of SNR was found ($p < 0.001$, $F_{(1,96)} = 20.02$, $\text{partial-}\eta^2 = 0.39$), with greater gain seen at lower SNRs. Similarly, a main effect of measure was seen ($p < 0.001$, $F_{(1,96)} = 152.51$, $\text{partial-}\eta^2 = 0.83$), with greater gain seen when calculating whole-word recognition. When collapsed across whole-word and phoneme recognition, no significant between-subjects effect of age was observed ($p = 0.63$, $F_{(1,96)} = 0.43$, $\text{partial-}\eta^2 = 0.02$). No significant 2-way interactions were found between SNR and measure ($p = 0.417$, $F_{(1,32)} = 0.68$, $\text{partial-}\eta^2 = 0.02$), between SNR and age ($p = 0.292$, $F_{(1,32)} = 1.15$, $\text{partial-}\eta^2 = 0.02$), or between age and measure ($p = 0.745$, $F_{(1,32)} = 0.11$, $\text{partial-}\eta^2 < 0.01$).

4. Discussion

One of the most common complaints of older adults is difficulty in understanding speech in noisy environments. In such environments seeing a speaker's mouth can have dramatic behavioral benefits (Sumbly and Pollack, 1954). Despite well-documented increases in multisensory integration associated with the pairing of

simple, non-speech stimuli in older adults (Freiherr et al., 2013; Laurienti et al., 2006), there have also been reports of decreases in behavioral gain associated with audiovisual speech integration in an aging cohort (Tye-Murray et al., 2010). The current results provide novel evidence to resolve this important conflict. Here, we show that older adults do in fact benefit from audiovisual speech integration in a similar manner to younger adults when tasked with phoneme recognition. However, at the level of whole-word recognition, older and younger adults showed different patterns of multisensory gain. Whereas younger adults' performance was consistent with the concept of inverse effectiveness (i.e., increasing gains in multisensory recognition as SNR decreased), older adults did not show more multisensory gain as SNR decreased beyond –6dB SPL.

Although changes in perception and cognition are frequently reported in normal aging, sensory processing also declines in healthy aging. These changes include a reduction in visual acuity and auditory sensitivity because of changes in transduction processes in the retina and cochlea (Baltes and Lindenberger, 1997; Lindenberger and Baltes, 1997; Lindenberger and Ghisletta, 2009), as well as age-related changes along the sensory processing hierarchies (Cliff et al., 2013; Hugenschmidt et al., 2009; Nagamatsu et al., 2011). In the present study, we saw such effects very clearly, with a significant main effect of age group for both the auditory-only and visual-only conditions in which younger adults performed better than older adults. Our interests extended beyond these changes in unisensory function to include how multisensory audiovisual abilities change with advancing age.

One hypothesis for the present study was that older individuals would benefit more from, or rely more on, multisensory integration to compensate for declines in unisensory acuity. This hypothesis is founded in the concept of inverse effectiveness, in which greater multisensory gains are seen as the information from the individual senses becomes weaker or more ambiguous (Laurienti et al., 2006; Peiffer et al., 2007; Stevenson et al., 2014a, 2014c). For example, despite being poorer lip readers (Sommers et al., 2005), older individuals attend to the speaker's mouth to a higher degree, presumably as a strategy to increase the available visual information. An alternative hypothesis contends that on reaching a threshold auditory SNR, there is a point at which individuals can no longer extract meaningful information from a given (or set of given) sensory input(s) and subsequently fail to exhibit any behavioral benefit from a second sensory input. For example, Gordon and Allen (2009) presented unisensory and multisensory sentences to both older and younger adults and varied the saliency of the visual speech signal. Older adults showed reduced performance under all conditions relative to younger individuals, but with non-degraded visual information, their multisensory gain remained similar. However, on degrading the visual input, older adults reached a critical threshold in which they failed to benefit from the visual signal (i.e., showed no multisensory gain), whereas the younger group continued to show improvements over the unisensory presentations. Although our present study did not actively manipulate the visual signal, the marginally worse performance of the older adults relative to the younger adults may have pushed these individuals closer to such a threshold.

The whole-word recognition data reported here do not unequivocally rule out one hypothesis over another. When looking at intermediate SNR levels (e.g., –6 dB), older adults show increased multisensory performance compared with younger adults, as measured by the difference between observed performance and predicted levels of accuracy based on pooled unisensory performance. This is consistent with prior work that has found the –6 dB SNR level to be a “sweet spot” for multisensory-mediated gains in speech intelligibility (Ma et al., 2009; Ross et al., 2007a, 2007b,

2011). The pattern of differences in gain between older and younger adults support the idea that older adults differentially benefit from multisensory integration relative to younger adults but only within a limited range of SNRs. For other SNRs, such as –18 dB, older adults showed substantially less gain when compared with younger adults. In contrast to the “inverted U” function seen for older adults with whole-word recognition, changes in multisensory gain as a function of SNR in younger adults for whole-word stimuli showed a declining monotonic function much more in keeping with inverse effectiveness. These data, taken in isolation, support the hypothesis by Gordon and Allen (2009) that when older adults reach a critical threshold, the benefit that can be gained from perceiving speech through multiple sensory modalities is reduced.

Although the whole-word recognition data suggest that there is an age-related decrease in multisensory gain at lower SNRs, phoneme recognition shows a different pattern. In contrast to whole-word recognition, there was no main effect of age difference in multisensory gain during phoneme recognition. In fact, despite significantly poorer absolute levels of performance across SNRs and sensory modalities relative to younger adults, older adults showed strikingly similar levels of multisensory gain, even down to the lowest SNR levels. This finding suggests that older adults are integrating the basic auditory (i.e., phonemes) and visual (i.e., visemes) building blocks of speech information similarly to younger individuals yet are failing to reach a critical information threshold at which they can correctly identify the whole word. One possible cause of these seemingly conflicting results is rooted in the behavioral measures used. During whole-word recognition, the listeners may still be integrating the phonemic and visemic information, yet fall short of recognizing the entire word. For example, a listener may be presented with the word “cat” in a unisensory (i.e., auditory alone, visual alone) context and not perceive any portion of the word correctly. However, when presented with the word in an audiovisual context, the same listener may report perceiving “cab.” Although “cab” is incorrect on the word level, such a response still represents an increase in the information perceived, because the participant was able to correctly perceive the initial 2 phonemes/visemes. Thus, scoring whole-word recognition may miss important elements of the perceptual improvement, particularly at low SNRs in which it is more difficult to correctly identify entire words.

These results underscore the importance of how a given response is measured when characterizing the presence and magnitude of multisensory integration (Mégevand et al., 2013; Stevenson and Wallace, 2013). Analogous results have been seen in the neurophysiological realm, in which early studies of multisensory integration in individual neurons were based on spiking (i.e., action potential) responses (Meredith and Stein, 1983, 1985). In the absence of clear changes in spiking under multisensory conditions, a neuron was said to not integrate its different inputs. However, more recent work, which has focused not only on spiking but also on changes in the local field potential (LFP, a measure of summed synaptic activity) has illustrated the presence of multisensory interactions in the LFP response in the absence of clear changes in neuronal spiking (Ghazanfar and Maier, 2009; Ghazanfar et al., 2005; Ghose and Wallace, 2014; Kayser et al., 2009; Sarko et al., 2013).

The dissociation between whole-word and phoneme-recognition performance in the present study illustrates that, at least for older adults, multisensory-mediated gains in intelligibility at the phonemic level do not necessarily translate to comparable gains at the whole-word level (as they appear to in the younger cohort). This finding suggests that additional processes, most notably those involved in the transformation of phonemic representations into word-based representations, may be preferentially

impacted in the aging process. Weaknesses in this transformation process may be reflective of more generalized cognitive declines that impact domains such as memory, executive function, and attention (for review, see [Freiherr et al., 2013](#)), each of which likely contribute to the assembly of larger functional speech units (e.g., words, sentences, and so forth). Conversely, changes in these cognitive processes may be, at least in part, because of less efficient processing within sensory and multisensory representations, given that these representations serve as the foundation for perceptual and cognitive representations.

The concept of lower-level sensory processing changes in healthy aging cascading into higher-level perceptual and cognitive difficulties is not unique to speech perception. Indeed, other areas of common complaint in older adults, such as memory, can be at least in part the result of changes in sensory processing ([Baltes and Lindenberger, 1997](#); [Burke et al., 2012](#); [Lindenberger and Ghisletta, 2009](#); [Lovden et al., 2005](#)). For example, in an analogous finding from the visual domain to that reported here, older adults performed equivalently to younger adults on visual perceptual discriminations requiring processing of only a single feature (such as color). However, when the visual discrimination task required binding, or integrating, multiple features to create a cohesive representation of the whole object, older adults were impaired ([Ryan et al., 2012](#)). Indeed, it has been argued that deficits in apparently distinct memory and perceptual functions may in fact arise from common representations and computational mechanisms ([Barense et al., 2012](#)). To inform these questions, future work should focus on establishing the nature of the relationships between (multi)sensory and cognitive representations, as well as describing how these relationships change with both normal and pathologic aging.

Despite the different patterns of multisensory gain across SNRs at the whole word and phoneme levels, it can be argued that whole-word recognition is a more relevant and meaningful measure of speech perception. Thus, in a typical speech comprehension setting, successful recognition at the word level will produce a significant behavioral benefit, whereas increased information processing at the phonemic level may be, in isolation, less behaviorally consequential. Although these results suggest that in terms of the integration of the basic building blocks for language (i.e., phonemic/visemic), inverse effectiveness appears to govern performance, from a more practical perspective there appears to be a “sweet spot” at intermediate SNRs at which behaviorally relevant levels of multisensory gain are greatest, at least in aging populations.

Although this study examined age-related changes in multisensory perception of speech across SNR levels, the effectiveness of a speech signal is only one of a number of factors influencing sensory integration. In addition to effectiveness, the temporal and spatial relationship of sensory signals are also very important. The more spatially congruent and temporally aligned two inputs are, the more likely they will be integrated ([Bertelson and Radeau, 1981](#); [Conrey and Pisoni, 2006](#); [Dixon and Spitz, 1980](#); [Massaro et al., 1996](#); [van Atteveldt et al., 2007](#); [van Wassenhove et al., 2007](#)). Thus, changes in spatial and/or temporal processing seen in healthy-aging populations may also impact audiovisual speech perception ([Hay-McCutcheon et al., 2009](#)). This is perhaps most germane with respect to temporal processing, as higher-acuity multisensory temporal processing has been linked to increases in integration of audiovisual speech in healthy individuals ([Stevenson et al., 2012a, 2012b](#)). Furthermore, clinical groups that show impaired multisensory temporal processing also show decreased integration of audiovisual speech ([Stevenson et al., 2014b](#)). This, coupled with strong evidence that declines in auditory temporal processing are related to auditory speech perception in older adults ([Gordon-Salant and Fitzgibbons, 1993](#); [Pichora-Fuller, 2003](#); [Schneider and Pichora-Fuller, 2001](#)), suggests that possible age-

related declines in multisensory temporal processing may likewise be related to age-related declines in audiovisual speech perception abilities.

5. Conclusion

Here, we reported results from an audiovisual speech recognition task using varying levels of noise with younger and older adult participants. We found that for whole-word recognition, older adults showed greater multisensory gains at intermediate SNRs compared with younger adults. On the other hand, at the phoneme level a different pattern emerged whereby both younger and older adults showed similar increases in multisensory gain as SNR decreased, consistent with the concept of inverse effectiveness. Collectively, the results provide important insights into both the similarities and differences in how older and younger adults integrate auditory and visual speech cues in noisy environments, and help explain some of the conflicting findings from previous studies of multisensory speech perception in healthy aging.

Disclosure statement

The authors have no actual or potential conflicts of interest.

Acknowledgements

Funding for this work was provided by a Banting Postdoctoral Fellowship administered by the Government of Canada, a University of Toronto Department of Psychology Postdoctoral Fellowship Grant, the Autism Research Training Program, National Institutes of Health F32 DC011993, R34 DC010927, and R21 CA1834892, a Simons Foundation research grant, a Vanderbilt Institute for Clinical and Translational Research grant VR7263, a Vanderbilt Kennedy Center MARI/Hobbs Award, the Vanderbilt Brain Institute, and the Vanderbilt University Kennedy Center.

References

- Baltes, P.B., Lindenberger, U., 1997. Emergence of a powerful connection between sensory and cognitive functions across the adult life span: a new window to the study of cognitive aging? *Psychol. Aging* 12, 12–21.
- Barense, M.D., Groen, L.I., Lee, A.C., Yeung, L.-K., Brady, S.M., Gregori, M., Kapur, N., Bussey, T.J., Saksida, L.M., Henson, R.N., 2012. Intact memory for irrelevant information impairs perception in amnesia. *Neuron* 75, 157–167.
- Bertelson, P., Radeau, M., 1981. Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept. Psychophys.* 29, 578–584.
- Bishop, C.W., Miller, L.M., 2009. A multisensory cortical network for understanding speech in noise. *J. Cogn. Neurosci.* 21, 1790–1805.
- Brainard, D.H., 1997. The psychophysics toolbox. *Spat. vis.* 10, 433–436.
- Burke, S.N., Ryan, L., Barnes, C.A., 2012. Characterizing cognitive aging of recognition memory and related processes in animal models and in humans. *Front. Aging Neurosci.* 4, 15.
- Cliff, M., Joyce, D.W., Lamar, M., Dannhauser, T., Tracy, D.K., Shergill, S.S., 2013. Aging effects on functional auditory and visual processing using fMRI with variable sensory loading. *Cortex* 49, 1304–1313.
- Conrey, B., Pisoni, D.B., 2006. Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *J. Acoust. Soc. Am.* 119, 4065–4073.
- Dixon, N.F., Spitz, L., 1980. The detection of auditory visual desynchrony. *Perception* 9, 719–721.
- Dubno, J.R., Dirks, D.D., Morgan, D.E., 1984. Effects of age and mild hearing loss on speech recognition in noise. *J. Acoust. Soc. Am.* 76, 87–96.
- Fraser, S., Gagne, J.P., Alepins, M., Dubois, P., 2010. Evaluating the effort expended to understand speech in noise using a dual-task paradigm: the effects of providing visual speech cues. *J. speech Lang. Hear. Res.* 53, 18–33.
- Freiherr, J., Lundström, J.N., Habel, U., Reetz, K., 2013. Multisensory integration mechanisms during aging. *Front. Hum. Neurosci.* 7, 863.
- Ghazanfar, A.A., Maier, J.X., 2009. Rhesus monkeys (*Macaca mulatta*) hear rising frequency sounds as looming. *Behav. Neurosci.* 123, 822–827.
- Ghazanfar, A.A., Maier, J.X., Hoffman, K.L., Logothetis, N.K., 2005. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J. Neurosci.* 25, 5004–5012.

- Ghose, D., Wallace, M., 2014. Heterogeneity in the spatial receptive field architecture of multisensory neurons of the superior colliculus and its effects on multisensory integration. *Neuroscience* 256, 147–162.
- Gordon, M.S., Allen, S., 2009. Audiovisual speech in older and younger adults: integrating a distorted visual signal with speech in noise. *Exp. Aging Res.* 35, 202–219.
- Gordon-Salant, S., Fitzgibbons, P.J., 1993. Temporal factors and speech recognition performance in young and elderly listeners. *J. Speech Lang. Hear. Res.* 36, 1276–1285.
- Gosselin, P.A., Gagne, J.P., 2011. Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. *Int. J. Audiol.* 50, 786–792.
- Hay-McCutcheon, M.J., Pisoni, D.B., Hunt, K.K., 2009. Audiovisual asynchrony detection and speech perception in hearing-impaired listeners with cochlear implants: a preliminary analysis. *Int. J. Audiol.* 48, 321–333.
- Hugenschmidt, C.E., Mozolic, J.L., Tan, H., Kraft, R.A., Laurienti, P.J., 2009. Age-related increase in cross-sensory noise in resting and steady-state cerebral perfusion. *Brain Topogr.* 21, 241–251.
- Humes, L.E., 1996. Speech understanding in the elderly. *J. Am. Acad. Audiol.* 7, 161–167.
- James, T.W., Stevenson, R.A., Kim, S., 2012. Inverse effectiveness in multisensory processing. In: Stein, B.E. (Ed.), *The New Handbook of Multisensory Processes*. MIT Press, Cambridge, MA.
- Kayser, C., Petkov, C.I., Logothetis, N.K., 2009. Multisensory interactions in primate auditory cortex: fMRI and electrophysiology. *Hear. Res.* 258, 80–88.
- Lachs, L., Hernandez, L.R., 1998. Update: the Hoosier audiovisual multitalker database. *Research on Spoken Language Processing Progress Report No. 22*, 377–388.
- Laurienti, P.J., Burdette, J.H., Maldjian, J.A., Wallace, M.T., 2006. Enhanced multisensory integration in older adults. *Neurobiol. Aging* 27, 1155–1163.
- Lindenberger, U., Baltes, P.B., 1997. Intellectual functioning in old and very old age: cross-sectional results from the Berlin Aging Study. *Psychol. Aging* 12, 410–432.
- Lindenberger, U., Ghisletta, P., 2009. Cognitive and sensory declines in old age: gauging the evidence for a common cause. *Psychol. Aging* 24, 1–16.
- Liu, X.Z., Yan, D., 2007. Ageing and hearing loss. *J. Pathol.* 211, 188–197.
- Lovden, M., Ghisletta, P., Lindenberger, U., 2005. Social participation attenuates decline in perceptual speed in old and very old age. *Psychol. Aging* 20, 423–434.
- Luce, P.A., Pisoni, D.B., 1998. Recognizing spoken words: The neighborhood activation model. *Ear Hear* 19, 1.
- Ma, W.J., Zhou, X., Ross, L.A., Foxe, J.J., Parra, L.C., 2009. Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space. *PLoS One* 4, e4638.
- Martin, J.S., Jerger, J.F., 2005. Some effects of aging on central auditory processing. *J. Rehabil. Res. Dev.* 42 (4 Suppl 2), 25–44.
- Massaro, D.W., Cohen, M.M., Smelee, P.M., 1996. Perception of asynchronous and conflicting visual and auditory speech. *J. Acoust. Soc. Am.* 100, 1777–1786.
- Mégevand, P., Molholm, S., Nayak, A., Foxe, J.J., 2013. Recalibration of the multisensory temporal window of integration results from changing task demands. *PLoS One* 8, e71608.
- Meredith, M.A., Stein, B.E., 1983. Interactions among converging sensory inputs in the superior colliculus. *Science* 221, 389–391.
- Meredith, M.A., Stein, B.E., 1985. Descending efferents from the superior colliculus relay integrated multisensory information. *Science* 227, 657–659.
- Meredith, M.A., Stein, B.E., 1986. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J. Neurophysiol.* 56, 640–662.
- Nagamatsu, L.S., Carolan, P., Liu-Ambrose, T.Y., Handy, T.C., 2011. Age-related changes in the attentional control of visual cortex: a selective problem in the left visual hemifield. *Neuropsychologia* 49, 1670–1678.
- Nath, A.R., Beauchamp, M.S., 2011. Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *J. Neurosci.* 31, 1704–1714.
- Park, D.C., Lautenschlager, G., Hedden, T., Davidson, N.S., Smith, A.D., Smith, P.K., 2002. Models of visuospatial and verbal memory across the adult life span. *Psychol. Aging* 17, 299–320.
- Peiffer, A.M., Mozolic, J.L., Hugenschmidt, C.E., Laurienti, P.J., 2007. Age-related multisensory enhancement in a simple audiovisual detection task. *Neuroreport* 18, 1077–1081.
- Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat. vis.* 10, 437–442.
- Pichora-Fuller, M.K., 2003. Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension. *Int. J. Audiol.* 42, S59–S67.
- Raab, D.H., 1962. Statistical facilitation of simple reaction times. *Trans. N.Y. Acad. Sci.* 24, 574–590.
- Ross, L.A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., Foxe, J.J., 2011. The development of multisensory speech perception continues into the late childhood years. *Eur. J. Neurosci.* 33, 2329–2337.
- Ross, L.A., Saint-Amour, D., Leavitt, V.M., Javitt, D.C., Foxe, J.J., 2007a. Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb. Cortex* 17, 1147–1153.
- Ross, L.A., Saint-Amour, D., Leavitt, V.M., Molholm, S., Javitt, D.C., Foxe, J.J., 2007b. Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophr. Res.* 97, 173–183.
- Ryan, L., Cardoza, J., Barense, M., Kawa, K., Wallentin-Flores, J., Arnold, W., Alexander, G., 2012. Age-related impairment in a complex object discrimination task that engages perirhinal cortex. *Hippocampus* 22, 1978–1989.
- Sarko, D.K., Ghose, D., Wallace, M.T., 2013. Convergent approaches toward the study of multisensory perception. *Front. Syst. Neurosci.* 7, 81.
- Schneider, B.A., Pichora-Fuller, M.K., 2001. Age-related Changes in Temporal Processing: Implications for Speech Perception. *Seminars in hearing*, pp. 227–240. Copyright © 2001 by Thieme Medical Publishers, Inc., New York, NY.
- Sheffert, S.M., Lachs, L., Hernandez, L.R., 1996. The Hoosier audiovisual multitalker database. *Research on Spoken Language Processing Progress Report No. 21*, 578–583.
- Sommers, M.S., Tye-Murray, N., Spehar, B., 2005. Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear Hear* 26, 263–275.
- Stevenson, R.A., Bushmakina, M., Kim, S., Wallace, M., Puce, A., James, T., 2012a. Inverse effectiveness and multisensory interactions in visual event-related potentials with audiovisual speech. *Brain Topogr.* 25, 308–326.
- Stevenson, R.A., Geoghegan, M.L., James, T.W., 2007. Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects. *Exp. Brain Res.* 179, 85–95.
- Stevenson, R.A., Ghose, D., Fister, J.K., Sarko, D.K., Altieri, N.A., Nidiffer, A.R., Kurela, L.R., Siemann, J.K., James, T.W., Wallace, M.T., 2014a. Identifying and quantifying multisensory integration: a tutorial review. *Brain Topogr.* <http://dx.doi.org/10.1007/s10548-014-0365-7>.
- Stevenson, R.A., James, T.W., 2009. Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage* 44, 1210–1223.
- Stevenson, R.A., Kim, S., James, T.W., 2009. An additive-factors design to disambiguate neuronal and areal convergence: measuring multisensory interactions between audio, visual, and haptic sensory streams using fMRI. *Exp. Brain Res.* 198, 183–194.
- Stevenson, R.A., Siemann, J.K., Schneider, B.C., Eberly, H.E., Woynarowski, T.G., Camarata, S.M., Wallace, M.T., 2014b. Multisensory temporal integration in autism spectrum disorders. *J. Neurosci.* 34, 691–697.
- Stevenson, R.A., VanDerKlok, R.M., Pisoni, D.B., James, T.W., 2011. Discrete neural substrates underlie complementary audiovisual speech integration processes. *Neuroimage* 55, 1339–1345.
- Stevenson, R.A., Wallace, M.T., 2013. Multisensory temporal integration: task and stimulus dependencies. *Exp. Brain Res.* 227, 249–261.
- Stevenson, R.A., Wallace, M.T., Altieri, N., 2014c. The interaction between stimulus factors and cognitive factors during multisensory integration of audiovisual speech. *Front. Psychol.* 5, 352.
- Stevenson, R.A., Zemtsov, R.K., Wallace, M.T., 2012b. Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 1517–1529.
- Sumby, W.H., Pollack, I., 1954. Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215.
- Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J., Hale, S., 2010. Aging, audio-visual integration, and the principle of inverse effectiveness. *Ear Hear* 31, 636–644.
- van Atteveldt, N.M., Formisano, E., Blomert, L., Goebel, R., 2007. The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb. Cortex* 17, 962–974.
- van Wassenhove, V., Grant, K.W., Poeppel, D., 2007. Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* 45, 598–607.
- Wallace, M.T., Wilkinson, L.K., Stein, B.E., 1996. Representation and integration of multiple sensory inputs in primate superior colliculus. *J. Neurophysiol.* 76, 1246–1266.
- Werner, S., Noppeney, U., 2009. Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cereb. Cortex* 20, 1829–1842.